The Relative Ineffectiveness of Bibliographic Search Engines

IVAN VALIELA AND PAULINA MARTINETTO

The increasing number of scientific publications has made bibliographic search engines essential tools in all disciplines. These software-based devices, however, are far from perfect. Comparisons of software-based bibliographic search engines with complete lists of three authors' publications showed that reference citations were not generally available before 1970, and that the effectiveness of recovery was improving but was quite variable, yielding on average 36 percent of the publications. There was marked year-to-year inconsistency in the recovery of titles. The inconsistency could not be explained by differences in indexing due to journal reputation: there was no evident relationship between search effectiveness and journal impact factor, but the percentage of recovered citations was higher for indexed journals. Search engines are widely used in bibliographic searches performed for evaluating researchers, awarding promotions, or assessing journal performance. Given the ineffectiveness of search engines, their use in making such important personal and institutional decisions needs careful consideration.

Keywords: search engines, scientific publications, bibliographic searches

Complaints about the overwhelming plethora of scientific publications have been voiced for many decades. An eloquent image of the problem appeared as early as 1939, showing scientists covering the world with their words (figure 1). Today the scientific community is blanketed in a blizzard of publications. For instance, using just one search engine to find papers on salt marshes—a topic that might be considered reasonably narrow—yielded 4064 publications in total, and 1520 appearing in the last 10 years. We are facing a torrent of information of unprecedented dimensions.

One response to the avalanche of papers being made available has been the development of a number of bibliographic search engines. These software-based devices have become essential tools in scientists' often inadequate efforts both to keep up with the literature in their specialties and to review what has been published in the past so as to prevent repeating studies. In our own use of these engines, however, we have often noticed that references we in fact had in our reprint collections failed to appear in the lists furnished by search engines. The essential tool that researchers all now use seemed to have some degree of ineffectiveness.

Studies of the citation of references available in search engines suggest that journal papers and book chapters dominate search engine results (Goodrum et al. 2001), that no one search engine provides sufficient coverage of the published literature (Lawrence and Giles 1998), and that online availability increases citation frequency (Lawrence 2001). To our knowledge, the fundamental issue, that of the quantitative effectiveness of search engines in finding published works, has not been investigated. Here we inquire as to the degree of effectiveness of popular search engines in finding published papers.

Figure 1. An allegory in which scientists cover the world with the written word. Originally published by Wegman (1939); reprinted from Valiela (2001).

Ivan Valiela (e-mail: valiela@bu.edu) is a professor of biology at the Boston University Marine Program, Marine Biological Laboratory, Woods Hole, MA 02543. He is the author of Marine Ecological Processes, *a widely used textbook, and* Doing Science, *a guide to how scientists carry out their work. His research focuses on natural and human controls of the structure and function of coastal ecosystems. Paulina Martinetto (e-mail: pmartin@mdp.edu.ar) is a doctoral candidate in marine ecology at the Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, CC573 Correo Central B7600WAG, Mar del Plata, Argentina, and a visiting investigator at the Boston University Marine Program. One focus of her current research is land–estuary and benthic–pelagic couplings.*

We tested a number of search engines by comparing the lists provided by searches using the search engines with the complete publication lists of three reasonably active researchers. The three authors for whose work we searched (Ramón Margalef, Ivan Valiela, and Oscar Iribarne) studied a variety of aspects of the ecology of aquatic environments, and we used search engines appropriate for these topics. Our conclusions therefore are restricted to these subjects, but our findings may be quite general. We encourage readers to repeat our exercise with their own lists of publications; we believe their findings will corroborate ours.

The choice of search engines was determined by the subjects and by their availability in the Marine Biological Laboratory–Woods Hole Oceanographic Institution Library. Searches were conducted by author name. The choice of authors was dictated largely by the fact that we had their complete lists of publications at hand. Comparison of these three lists allowed us to assess the effectiveness of recovery of titles from authors at different stages of their careers, and from work spanning many decades (1940 onward; figure 2).

We found that references earlier than 1970 were generally not available in any of the search engines that we used (figure 2). This is reasonable enough, since higher priority has to be given to recent work in the entry of information. It is curious, however, that some pre-1970 entries do appear in some search engines, even though the text describing the software gives later start dates.

The percentage of papers found increased variably across the decades (figure 3). It is of some reassurance that the effectiveness of the search engines generally increased toward the end of the 20th century, but two features in our findings are surprising.

First, the recovery of publications was highly variable: most post-1970 points fell between 0 and 100 percent (figure 3). On average, considering papers by the three authors for post-1970 data, only 55 percent of the papers were found (table 1). Even for Iribarne—the youngest of the authors, whose publications appeared in recent years—only 78 percent of papers were found (table 1). For all years, we found that only 36 percent published by the three authors were found (table 1). Perhaps this number was affected by the fact that Margalef

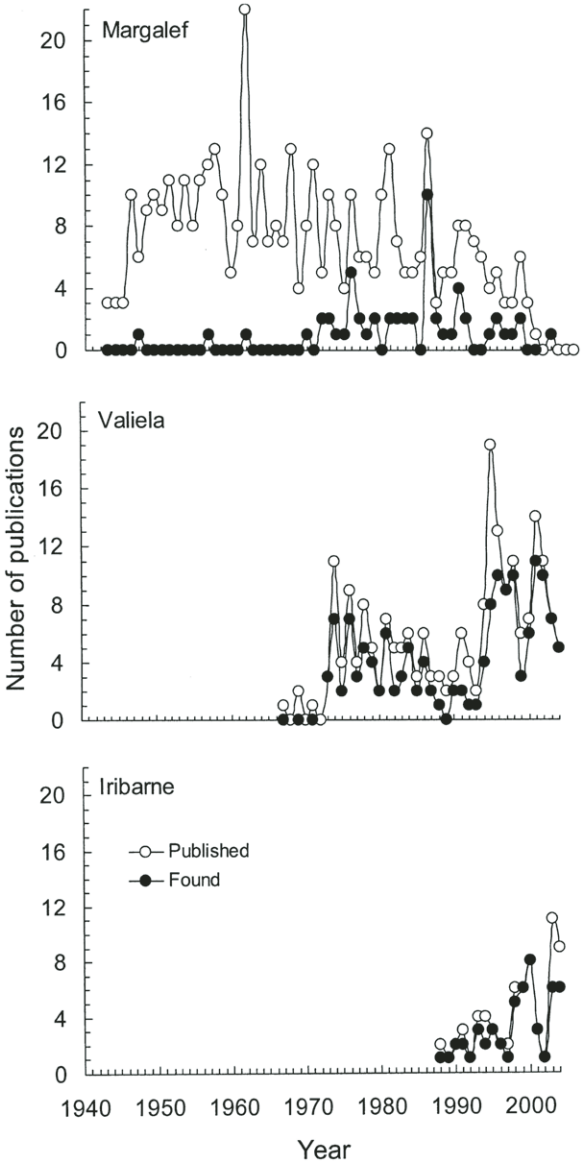

Figure 2. Number of publications published and found per year, using all databases taken together, for Ramón Margalef, Ivan Valiela, and Oscar Iribarne.

Table 1. Percentage of publications of Ramón Margalef, Ivan Valiela, and Oscar Iribarne found in the Aquatic Sciences and Fisheries Abstracts database, in the Biological Sciences database, and in a group of databases for the periods 1943–1970, 1971–2004, and 1943–2004.

	Percentage of publications found for the period								
	1943–1970			1971–2004			1943–2004		
Author	ASFA	Biological Sciences	Group	ASFA	Biological Sciences	Group	ASFA	Biological Sciences	Group
Margalef	1.2	1.2	1.6	20.3	21.4	28.6	9.3	9.7	13.0
Valiela	0	0	0	57.8	55.0	69.7	57.0	54.2	68.7
Iribarne	–	–	–	70.6	75.0	77.9	70.6	75.0	77.9
Total	1.2	1.2	1.6	44.8	44.4	54.8	29.4	29.1	35.9

Note: The group of databases, from Cambridge Scientific Abstracts, Internet Database Services, comprises Aquatic Sciences and Fisheries Abstracts (ASFA; 1971–current), Biological Sciences (1982–current), Biology Digest (1989–current), Conference Papers Index (1982–current), GeoRef (1785–current), MEDLINE (1993–current), Oceanic Abstracts (1981–current), Plant Science (1994–current), PsycINFO (1840–current), TOXLINE (1999–current), Water Resources (1967–current), and Abstracts Zoological Record Plus (2004–current).

Table 2. Number of years in which 100 percent, 50–99 percent, 1–49 percent, and 0 percent of the publications of Ramón Margalef, Ivan Valiela, Oscar Iribarne, and all three authors were found in the Aquatic Sciences and Fisheries Abstracts database, in the Biological Sciences database, and in a group of databases.

	Number of years with percentage search engine effectiveness of											
	100			50–99			1–49			0		
Author	ASFA	Biological Sciences	Group	ASFA	Biological Sciences	Group	ASFA	Biological Sciences	Group	ASFA	Biological Sciences	Group
Margalef	0	0	2	1	1	4	22	22	23	36	36	30
Valiela	3	2	5	17	19	21	10	9	5	5	5	4
Iribarne	9	9	9	6	7	8	2	1	0	0	0	0
Total	0	0	0	12	13	18	25	24	20	25	25	24

Note: The group of databases, from Cambridge Scientific Abstracts, Internet Database Services, comprises the Aquatic Sciences and Fisheries Abstracts (ASFA; 1971–current), Biological Sciences (1982–current), Biology Digest (1989–current), Conference Papers Index (1982–current), GeoRef (1785–current), MEDLINE (1993–current), Oceanic Abstracts (1981–current), Plant Science (1994–current), PsycINFO (1840–current), TOXLINE (1999–current), Water Resources (1967–current), and Abstracts Zoological Record Plus (2004–current).

published in Spanish and Catalan journals, which were not effectively indexed. Nonetheless, of the 53 of Margalef's 409 publications that appeared in international journals, only 13 were found (25 percent). This is quite low. Even considering publications by Valiela and Iribarne, who published in international journals, only 69 percent and 78 percent of the papers, respectively, were found by the search engines. These recovery rates for recent authors seem too low to be acceptable. Incidentally, the effectiveness of the different search engines was similar (table 1); for simplicity in figure 3, we reported the publications found using all the search engines.

Second, it was also surprising that the effectiveness of the search engines was so variable from year to year. For instance, only 50 to 60 percent of Iribarne's papers dated 2003 or 2004 were found, while 100 percent of those dated from 1999 to 2002 were found (figure 3). There was therefore a substantial inconsistency in year-to-year records; it is unclear how this lack of consistency arises. Furthermore, there were very few years in which 100 percent of the papers were found (table 2). Fortunately, for recent authors, such as Iribarne, there were no years in which none of the published papers were found (table 2).

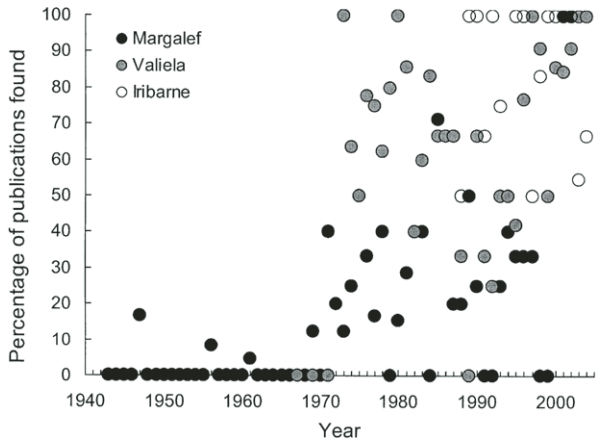

Figure 3. Percentage of publications found per year, using all databases together, for Ramón Margalef, Ivan Valiela, and Oscar Iribarne.

One possible explanation for the variation in search effectiveness might be that some journals (perhaps those with higher impact ranking) are better indexed by search engine staff than other journals. To examine that possibility, we compiled a list of the search engine effectiveness, and the ISI journal impact factor (Popescu 2002), for the journals in which our three authors published papers after 1970 (table 3).

There was no evident relationship between search effectiveness and journal impact factor (figure 4a). This was true even for journals in which the authors published five or more papers (which may have provided more reliable estimates of search efficiencies) (figure 4a). We note in passing that of two papers that appeared in the same year in the journal with the highest impact factor (27.9; table 3) (this journal was not included in figure 4), only one of the two was found by the search engines. In general, however, indexed journals had higher percentages of recovery than unindexed journals ($Z = 7.265$, $P < 0.001$). The recovery of papers that appeared in unindexed journals was low (figure 4b, inset). These results could account

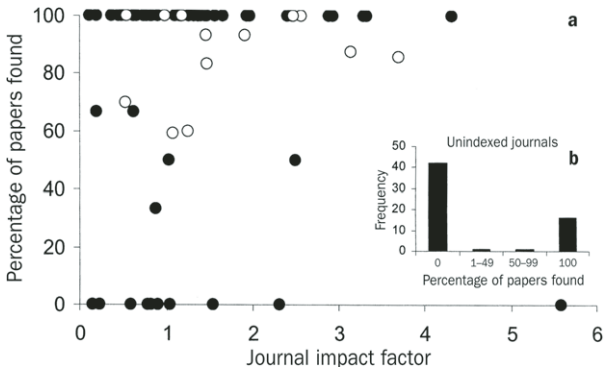

Figure 4. (a) Relationship between journal impact factor and the percentage of publications found using search engines for each journal. Open circles are journals in which the authors published 5 papers or more. (b) Frequency of unindexed journals with 0 percent, 1–49 percent, 50–99 percent, and 100 percent of papers found using search engines.

Table 3. Total number of papers published by Ramón Margalef, Ivan Valiela, and Oscar Iribarne between 1970 and 2004, and number of papers found using all databases together, sorted by journal.

Journal	Number of papers	Number of papers found	Efficiency	Journal impact factor
Indexed				
Ambio	1	1	1	1.4
American Journal of Botany	1	1	1	2.5
American Naturalist	2	2	1	4.3
American Scientist	1	0	0	1.0
Applied Geochemistry	1	1	1	1.3
Aquaculture	1	0	0	1.5
Aquatic Botany	1	1	1	1.4
Atmospheric Environment	1	0	0	2.3
Biogeochemistry	5	5	1	2.6
Biological Bulletin	49	29	0.6	1.1
BioScience	1	1	1	3.3
BioSystems	1	1	1	0.7
Botanica Marina	2	1	0.5	1.0
Bulletin of Marine Science	5	5	1	0.6
Canadian Entomologist	1	0	0	0.6
Canadian Journal of Fisheries and Aquatic Sciences	3	3	1	1.7
Canadian Journal of Zoology	1	1	1	1.2
Colonial Waterbirds	1	0	0	0.3[a]
Comparative Biochemistry and Physiology B	1	1	1	0.8
Earth Surface Processes and Landforms	1	0	0	1.0
Ecological Applications	2	2	1	3.3
Ecology	7	6	0.9	3.7
Ecology Letters	1	1	1	1.9
Ecosystems	1	1	1	2.4
Environmental Biology of Fishes	1	0	0	0.8
Environmental Management	2	0	0	0.8
Environmental Pollution	3	3	1	1.6
Environmental Toxicology and Chemistry	1	1	1	2.0
Estuaries	15	14	0.9	1.4
Estuarine Coastal and Marine Science	5	3	0.6	1.2[a]
Estuarine Coastal and Shelf Science	7	7	1	1.0
Fisheries Research	3	1	0.3	0.9
Hydrobiologia	2	2	1	0.7
Interciencia	1	1	1	0.2
International Journal of General Systems	1	0	0	0.2
Journal of Applied Ecology	3	3	1	2.9
Journal of Coastal Research	2	2	1	0.6
Journal of Environmental Quality	1	1	1	1.2
Journal of Experimental Marine Biology and Ecology	18	15	0.8	1.5
Journal of Marine Research	1	1	1	1.5
Journal of Plankton Research	1	1	1	1.3
Journal of Shellfish Research	10	7	0.7	0.5
Journal of the Marine Biological Association of the United Kingdom	1	1	1	0.8
Journal of Zoology	1	1	1	1.1
Lethaia	1	1	1	0.9
Limnology and Oceanography	16	14	0.9	3.1
Marine Biology	12	10	0.8	1.5
Marine Ecology	1	1	1	0.4
Marine Ecology Progress Series	15	14	0.9	1.9
Marine Environmental Research	1	1	1	0.9
Marine and Freshwater Behavior and Physiology	2	2	1	0.4
Marine Pollution Bulletin	6	6	1	1.2
Marine Technology Society Journal	1	0	0	0.1
Microbial Ecology	1	1	1	2.9
Nature	2	1	0.5	27.9
Oceanologica Acta	3	2	0.7	0.6
Oceanus	3	2	0.7	0.2
Oecologia	5	5	1	2.5
Oikos	2	1	0.5	2.5
Quarterly Review of Biology	3	0	0	5.6
Schweizerische Zeitschrift für Hydrologie	1	1	1	0.5[a]
Science of the Total Environment	1	1	1	1.4
Scientia Marina	5	5	1	0.5
Veliger	1	1	1	0.5
Water, Air, and Soil Pollution	2	2	1	0.8
Water Research	2	1	0.5	1.4
Not indexed				
Acta Botánica de Barcelona	1	0	0	
Acta Geológica Hispánica	1	1	1	
Anais da Academia Brasileira de Ciências	1	0	0	
Anales del Instituto Botánico AJ Cavanilles	1	0	0	
Annales de l'Institut Oceanographique, Paris, Nouvelle Serie	1	1	1	
Aquatic Ecology	2	2	1	
Arbor	1	0	0	
Arrels	1	0	0	
Avances del Saber	1	0	0	
Biological Invasions	1	1	1	
Boletín de la Estación Central de Ecología	1	1	1	
Boletín Informativo de la Fundación Juan March	1	0	0	
Boletín de la Sociedad de Historia Natural de Baleares	2	0	0	
Bulletí de la Societat Catalana Biología	2	0	0	
Canadian Bulletin of Fisheries and Aquatic Sciences	1	1	1	
Canadian Water Resources Journal	1	1	1	
Ciencia	1	0	0	
Ciencia Hoy	1	0	0	
Ciencies Experimentals i Matematiques	1	0	0	
Coastal and Estuarine Science	1	1	1	
Coenoses	1	1	1	
Collectanea Botanica	1	0	0	
Cuadernos da Área de Ciencias Mariñas, Seminario de Estudos Galegos	1	0	0	
Environmental Cape Cod	1	0	0	
EOS, Transactions of the American Geophysical Union	1	0	0	
Ethnica	1	0	0	
Gayana	1	0	0	
Graellsia	1	0	0	
Investigación y Ciencia	2	0	0	
Investigaciones Pesqueras	14	6	0.4	
Journal of Mediterranean Ecology	1	0	0	
Kobie	11	0	0	
La Recherche	1	0	0	
Malacological Review	1	0	0	
Memorias de la Real Academia de Ciencias y Artes de Barcelona	4	0	0	
Memorias de la Sociedad de Ciencias Naturales de La Salle	1	0	0	
Memorie dell'Instituto Italiano di Idrobiologia	1	1	1	
Microbiología	1	1	1	
Monografies de l'Equipe	1	0	0	
Mundo Científico	1	0	0	
Munibe	1	0	0	
Naturalia Hispánica	1	0	0	
Northwest Environmental Journal	1	0	0	
Oecologia Aquatica	4	3	0.8	
Paleontología y Evolución	2	2	1	
Perspectiva en Oceanografía	1	0	0	
Perspectiva Social	1	0	0	
Pirineos	1	0	0	
Quaderns d'Ecologia Aplicada	1	0	0	
Rapports de la Commission Internationale pour l'Exploration Scientifique de la Mer Méditerranée	1	0	0	
Real Academia de Ciencias Exactas, Físicas y Naturales de Madrid	1	0	0	
Revista de la Asociación de Ciencias Naturales del Litoral	1	1	1	
The Siren	1	0	0	
Spelean	1	0	0	
Studia Oecologica	1	1	1	
Tethys	1	0	0	
Treballs de la Societat Catalana de Biologia	3	0	0	
UNESCO, Ciencias Marinas	1	0	0	
Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie	1	1	1	

Note: Unless otherwise noted, journal impact factors are from ISI (2002). Efficiency was calculated by dividing the number of papers found by the number of papers published.
a. Average journal impact factors from ISI reports between 1974 and 2000 are from Popescu (2002).

for the low percentage of recovery for Margalef's publications, but there is still no ready explanation for the considerable disparities in search engine effectiveness that we found among journals and years for the papers published by the other two authors.

Newer search engines were introduced as this article was being written. For example, Scholar Google (*http://scholar.google.com*), a search engine available to users in a testing stage, found 18.7 percent of the papers for Margalef, 66.5 percent for Valiela, and 75.0 percent for Iribarne. These results corroborate our finding that today's search technology remains uncertain. Indeed, although Scholar Google produced a higher percentage recovery of papers published between 1943 and 1970 (15.0 percent) than did ASFA and Biological Sciences (1.6 percent), the percentage for papers published between 1971 and 2004 was lower (51.3 percent compared with 54.8 percent).

Given the number of papers being published, researchers must use search engines, and will increasingly do so in the future. The results just described, however, show that this essential tool is far from perfect. The designers of search engines need to improve their current data gathering considerably, with more balanced coverage of different journals, considerably more even coverage from one year to the next, and expanded coverage of past publications. Such improvements are critical, for many reasons beyond the need for scientists to keep up with advances in their specialties.

Citation searches using online engines are frequently used—properly or not—as measures of academic success, to assign positions, assess progress, decide on promotions, and inform other important life decisions. The journals that carry researchers' published work are also judged by their impact factor, and the output of departments and whole institutions is assessed by means of citation indexes derived from search engine surveys. It is clear that the impact of search engines has profound dimensions in personal and institutional decisions.

As libraries face the challenges of increased subscription costs and diminishing physical space, the scientific community will depend more and more on electronic means of access to its literature. As research funds become increasingly scarcer, it is also imperative to avoid "rediscovering wheels," even as the number of publications explodes. For these and more reasons, we will need improved effectiveness of search engines, which will be scientists' only way to keep contact with the past history of research and the widening of new developments in our field of study.

There is no doubt that scientists' use of software search engines will grow. While appreciating the essential function of these engines, it seems salutary to be aware of the imperfection of the data records they produce, and take measures to supplement their use with more traditional methods for bibliographic searches.

Acknowledgments

We are particularly grateful to Oscar Iribarne for providing his complete list of publications and for comments that helped to improve our manuscript, and to Josep Maria Gili for providing the list of publications of the late Ramón Margalef. We also thank Joanna York, Jennifer Bowen, and Mirta Teichberg for useful comments on an earlier draft and help with the graphics. Wayne Linklater and two anonymous reviewers provided useful suggestions.

References cited

Goodrum AA, McCain KW, Lawrence S, Giles CL. 2001. Computer Science Literature and the World Wide Web. (20 May 2005; *http://citeseer.ist.psu.edu/goodrum01computer.html*)

ISI. 2002. SCI Journal Citation Report. Philadelphia: Thomson Institute for Scientific Information.

Lawrence S. 2001. Online or invisible? Nature 411: 521.

Lawrence S, Giles CL. 1998. Searching the World Wide Web. Science 280: 98–100.

Popescu II. 2002. Science Journal Ranking by Average Impact Factors, Version 2002. (4 July 2005; *www.utdallas.edu/research/quantum/JournalRanking/index.html*)

Valiela I. 2001. Doing Science: Design, Analysis, and Communications of Scientific Research. New York: Oxford University Press.

Wegman CE. 1939. Zwei Bilder für das Arbeitszimmer eines Geologen. Geologische Rundschau 30: 1–392.